2023 年 9 月 重庆师范大学学报(自然科学版) Sep. 2023
第 40 卷 第 5 期 Journal of Chongqing Normal University (Natural Science) Vol. 40 No. 5

运筹学与控制论

DOI:10.11721/cqnuj20230518

求解非凸两分块优化问题的 Majorized Bregman 交替方向乘子法*

陈建华，彭建文，罗洪林
(重庆师范大学 数学科学学院，重庆 401331)

摘要:针对一类两分块非凸优化问题,提出 Majorized 带 Bregman 距离的交替方向乘子法。为了使问题的子问题更易求解,对目标函数中的光滑项进行极大化线性处理,并对 x 子问题和 y 子问题同时添加一个 Bregman 距离。在适当的假设条件下,建立了算法的全局收敛性。同时,在效益函数满足 KL 性质时,建立了算法的强收敛性。数值实验结果验证该算法的有效性。

关键词:交替方向乘子法;Bregman 距离;非凸优化问题;KL 性质;收敛性

中图分类号:O211.6 **文献标志码**:A **文章编号**:1672-6693(2023)05-0001-10

本文考虑如下带线性等式约束的两块可分非凸优化问题:

$$\begin{aligned} &\min\ f(\boldsymbol{x})+g(\boldsymbol{y}),\\ &\text{s.t.}\ \boldsymbol{Ax}-\boldsymbol{By}=\boldsymbol{0}。\end{aligned} \tag{1}$$

式中 $f:\mathbf{R}^n\to\mathbf{R}$ 是梯度 Lipschitz 连续的凸函数,$g:\mathbf{R}^m\to\mathbf{R}\cup\{+\infty\}$ 是正常下半连续函数,$\boldsymbol{A}\in\mathbf{R}^{l\times n}$,$\boldsymbol{B}\in\mathbf{R}^{l\times m}$。问题(1)在机器学习、信号处理、统计估计等方面有许多应用,其中 f 通常表示损失函数,g 表示正则化项,比如 g 可以是 l_q 范数($0<q\leqslant 1$)的平滑裁剪绝对偏差惩罚函数或极小极大凹惩罚函数,关于问题(1)的更多应用见文献[1]。

经典的交替方向乘子法(ADMM)是解决问题(1)的一种算法,它最初是由文献[2-3]提出。求解问题(1)的经典 ADMM 的迭代格式如下:

$$\begin{cases} \boldsymbol{y}^{k+1}=\arg\min\limits_{\boldsymbol{y}} L_\beta(\boldsymbol{x}^k,\boldsymbol{y},\boldsymbol{\lambda}^k),\\ \boldsymbol{x}^{k+1}=\arg\min\limits_{\boldsymbol{x}} L_\beta(\boldsymbol{x},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k),\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k+\beta(\boldsymbol{Ax}^{k+1}-\boldsymbol{By}^{k+1})。\end{cases} \tag{2}$$

其中 $L_\beta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})=f(\boldsymbol{x})+g(\boldsymbol{y})+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{Ax}-\boldsymbol{By})+\dfrac{\beta}{2}\|\boldsymbol{Ax}-\boldsymbol{By}\|_2^2$ 是问题(1)的增广拉格朗日函数。这里 $\boldsymbol{\lambda}\in\mathbf{R}^l$ 为拉格朗日乘子,$\beta>0$ 是一个惩罚参数。当 f 和 g 是凸函数时,ADMM 已被广泛研究并用于求解线性约束凸优化问题。

近些年,有关 ADMM 及它的变体在非凸环境下的收敛性分析的研究不断涌现,非凸环境即问题(1)中 f 和 g 至少有一个可能是非凸的。例如,Li 等人[4]研究了问题(1)的一种特殊形式的 ADMM,在目标函数满足 KL 性质和惩罚参数足够大的假设下建立了整个序列的全局收敛性。在惩罚参数的类似假设下,Hong 等人[5]分析了 ADMM 在解决某些非凸一致性和共享问题时的收敛性。Guo 等人[6]在增广拉格朗日函数满足 KL 性质且惩罚参数大于 $2L$ 的假设下,证明了由 ADMM 生成的迭代序列收敛到所考虑问题的临界点,其中 L 表示目标函数中

* **收稿日期**:2022-01-21 **修回日期**:2022-07-28 **网络出版时间**:2023-10-30T17:05

资助项目:国家自然科学基金重大项目(No. 11991024);国家自然科学基金面上项目(No. 12271071);重庆英才创新创业领军人才创新创业示范团队项目(No. CQYC20210309536);重庆英才计划“包干制”项目(No. cstc2022ycjh-bgzxm0147);重庆市高校创新研究群体项目(No. CXQT20014);重庆市自然科学基金项目(No. cstc2021jcyj-msxmX0300)

第一作者简介:陈建华,男,研究方向为最优化理论与算法,E-mail:1224888402@qq.com;**通信作者**:彭建文,男,教授,博士生导师,E-mail:jwpeng168@hotmail.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20231027.1738.007

光滑函数的梯度 Lipschitz 常数。最近,Wang 等人[7]在目标函数、涉及矩阵和惩罚参数满足一些常见的条件下,建立了对于问题(1)的经典 ADMM 的全局收敛性。

但是,通常式(2)中至少有 1 个子问题在实际应用中难以解决。为了克服这一缺点,文献[8-9]讨论了一个有效的变体是邻近 ADMM,尤其是针对凸优化问题。邻近 ADMM 引入了一些邻近项,使式(2)的子问题易于求解。针对非凸问题,Wang 等人[10]对经典 ADMM 的两个子问题都附以 Bregman 距离函数,在要求 Bregman 距离函数强凸的情况下,建立了 BADMM 的收敛性结果。针对问题(1)的 BADMM 算法的迭代格式如式(3)所示:

$$\begin{cases}\boldsymbol{y}^{k+1}=\arg\min\limits_{\boldsymbol{y}} L_\beta(\boldsymbol{x}^k,\boldsymbol{y},\boldsymbol{\lambda}^k)+\Delta_\psi(\boldsymbol{y},\boldsymbol{y}^k),\\ \boldsymbol{x}^{k+1}=\arg\min\limits_{\boldsymbol{x}} L_\beta(\boldsymbol{x},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k)+\Delta_\xi(\boldsymbol{x},\boldsymbol{x}^k),\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k+\beta(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})。\end{cases} \tag{3}$$

式中:Δ_ψ 和 Δ_ξ 分别表示关于函数 ψ 和 ξ 的 Bregman 距离。因为目标函数中的 f 是光滑凸函数,所以存在半正定矩阵 $\boldsymbol{\Sigma}_f$ 使得对任意的 $\boldsymbol{x},\boldsymbol{x}'\in\mathbf{R}^n$,都有:$f(\boldsymbol{x})\geqslant f(\boldsymbol{x}')+\langle\nabla f(\boldsymbol{x}'),\boldsymbol{x}-\boldsymbol{x}'\rangle+\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}'\|^2_{\boldsymbol{\Sigma}_f}$,又因为 f 是梯度 Lipschitz 连续的,所以存在半正定矩阵 $\hat{\boldsymbol{\Sigma}}_f$,且 $\hat{\boldsymbol{\Sigma}}_f\succeq\boldsymbol{\Sigma}_f$,使得对任意的 $\boldsymbol{x},\boldsymbol{x}'\in\mathbf{R}^n$,都有:

$$f(\boldsymbol{x})\leqslant\hat{f}(\boldsymbol{x},\boldsymbol{x}'):=f(\boldsymbol{x}')+\langle\nabla f(\boldsymbol{x}'),\boldsymbol{x}-\boldsymbol{x}'\rangle+\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}'\|^2_{\hat{\boldsymbol{\Sigma}}_f}。\tag{4}$$

关于问题(1)的 Majorized 增广拉格朗日函数定义为:

$$\hat{L}_\beta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{x}')=\hat{f}(\boldsymbol{x},\boldsymbol{x}')+g(\boldsymbol{y})+\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{B}\boldsymbol{y})+\frac{\beta}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{B}\boldsymbol{y}\|^2,$$

其中 $\hat{f}$ 的定义为式(4)。

因此本文提出的 Majorized BADMM 算法对于解决问题(1)的算法框架如下。

算法 1 步骤 1,令 $\beta>0$ 为给定的参数,Δ_ψ 和 Δ_ξ 为关于函数 ψ 和 ξ 的 Bregman 距离;

步骤 2,$\boldsymbol{y}^{k+1}=\arg\min\limits_{\boldsymbol{y}}\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)+\Delta_\psi(\boldsymbol{y},\boldsymbol{y}^k)$;

步骤 3,$\boldsymbol{x}^{k+1}=\arg\min\limits_{\boldsymbol{x}}\hat{L}_\beta(\boldsymbol{x},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)+\Delta_\xi(\boldsymbol{x},\boldsymbol{x}^k)$;

步骤 4,$\boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k+\beta(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})$;

步骤 5,如果终止条件未满足,令 $k=k+1$,并执行步骤 1。

本文提出的 Majorized BADMM 结合了 Bregman ADMM 和 Majorized ADMM 的优点,首先通过对目标函数中梯度 Lipschitz 连续的凸函数进行极大化线性处理,简化了关于该函数的子问题的求解;其次,通过选取合适的 Bregman 距离简化了子问题的求解,且加入 Bregman 距离后,建立算法 Majorized BADMM 的全局收敛性不需要目标函数的强凸性。本文将极大化线性技术用于处理非凸优化问题,最后将算法应用到稀疏逻辑回归问题中,通过与式(2)的数值比较,发现 2 种算法都可以在几乎相同的目标函数值下实现最优解,并且 Majorized BADMM 比 ADMM 花费的计算时间更少,且随着训练集维度的增大,优势更加突出,体现了算法 Majorized BADMM 的优越性。

1 预备知识

本文用 $\mathbf{R}^n$ 表示 n 维欧式空间,$\|\cdot\|$ 表示欧式范数,对任意 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$,规定它们的内积为 $\langle\boldsymbol{x},\boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$,规定 $\boldsymbol{x}$ 的 $\boldsymbol{G}$ 范数 $\|\boldsymbol{x}\|_G=\sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x}}$,这里 $\boldsymbol{G}$ 为半正定矩阵。$\boldsymbol{G}\succeq(\succ)0$ 说明 $\boldsymbol{G}$ 是半正定(正定)矩阵,$\lambda_{\min}(\boldsymbol{G})$ 与 $\lambda_{\max}(\boldsymbol{G})$ 分别表示半正定矩阵 $\boldsymbol{G}$ 的最小特征值与最大特征值。

对任意集合 $S\subseteq\mathbf{R}^n$,任意点 $\boldsymbol{x}\in\mathbf{R}^n$ 到 S 的距离定义如下:

$$\operatorname{dist}(\boldsymbol{x},S)=\begin{cases}\inf\limits_{y\in S}\|\boldsymbol{y}-\boldsymbol{x}\|,S\neq\varnothing,\\ +\infty,S=\varnothing。\end{cases}$$

定义函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 的有效域为:$\operatorname{dom} f=\{\boldsymbol{x}\in\mathbf{R}^n:f(\boldsymbol{x})<+\infty\}$,若 $\operatorname{dom} f\neq\varnothing$,则称 f 为正常函数。

定义 1[13] 若函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 在 $\boldsymbol{x}_0$ 处满足 $f(\boldsymbol{x}_0)\leqslant\liminf\limits_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$,则表示函数 f 在 $\boldsymbol{x}_0$ 处是下半连

续的，假设 f 在有效域内的每一点都满足下半连续性，则称 f 为下半连续函数。

定义 2[13]　设函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 是正常下半连续函数。

1）函数 f 在 $\boldsymbol{x}\in\mathrm{dom}\,f$ 处的 Fréchet 次微分，定义为所有满足下面条件的 $\boldsymbol{x}^*$ 的集合，记作 $\hat{\partial}f(\boldsymbol{x})$，

$$\liminf_{\boldsymbol{y}\to\boldsymbol{x},\boldsymbol{y}\to\boldsymbol{x}}\frac{f(\boldsymbol{y})-f(\boldsymbol{x})-\langle\boldsymbol{x}^*,\boldsymbol{y}-\boldsymbol{x}\rangle}{\|\boldsymbol{y}-\boldsymbol{x}\|}\geqslant 0,$$

若 $\boldsymbol{x}\notin\mathrm{dom}\,f$，令 $\hat{\partial}f(\boldsymbol{x})=\varnothing$。

2）函数 f 在 $\boldsymbol{x}\in\mathrm{dom}\,f$ 处的极限次微分，记为 $\partial f(\boldsymbol{x})$，定义如下：

$$\partial f(\boldsymbol{x})=\{\boldsymbol{x}^*\in\mathbf{R}^n:\exists\,\boldsymbol{x}_n\to\boldsymbol{x},f(\boldsymbol{x}_n)\to f(\boldsymbol{x}),\boldsymbol{x}_n^*\in\hat{\partial}f(\boldsymbol{x}_n),\boldsymbol{x}_n^*\to\boldsymbol{x}^*\}。$$

命题 1　从上述定义可以得到如下次微分的性质：

1）对 $\forall\,\boldsymbol{x}\in\mathbf{R}^n$，$\hat{\partial}f(\boldsymbol{x})\subseteq\partial f(\boldsymbol{x})$，其中 $\hat{\partial}f(\boldsymbol{x})$ 为闭凸集，而 $\partial f(\boldsymbol{x})$ 仅为闭集；

2）设 $\boldsymbol{x}_k^*\in\partial f(\boldsymbol{x}_k)$ 且 $\lim\limits_{k\to+\infty}(\boldsymbol{x}_k,\boldsymbol{x}_k^*)=(\boldsymbol{x},\boldsymbol{x}^*)$，则 $\boldsymbol{x}^*\in\partial f(\boldsymbol{x})$；

3）若 $\boldsymbol{x}\in\mathbf{R}^n$ 为 f 的极小点，则 $0\in\partial f(\boldsymbol{x})$。若 $0\in\partial f(\boldsymbol{x})$，则称 $\boldsymbol{x}$ 为 f 的稳定点（或临界点），函数 f 稳定点的集合记为 crit f；

4）若函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 为正常下半连续函数，$g:\mathbf{R}^n\to\mathbf{R}$ 连续可微，则对于任意 $\boldsymbol{x}\in\mathrm{dom}\,f$，有：

$$\partial(f+g)(\boldsymbol{x})=\partial f(\boldsymbol{x})+\nabla g(\boldsymbol{x})。$$

定义 3　若 $\boldsymbol{z}^*:=(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*)$ 为问题(1.1)的增广拉格朗日函数 $L_\beta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})$ 的稳定点，即 $0\in\partial L_\beta(\boldsymbol{z}^*)$，当且仅当：

$$\begin{cases}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^*=-\nabla f(\boldsymbol{x}^*),\\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\lambda}^*\in\partial g(\boldsymbol{y}^*),\\ \boldsymbol{A}\boldsymbol{x}^*-\boldsymbol{B}\boldsymbol{y}^*=0。\end{cases}$$

定义 4[13]　函数 f 称为 l_f-Lipschitz 连续，若对任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}\,f$ 都有 $\|f(\boldsymbol{x})-f(\boldsymbol{y})\|\leqslant l_f\|\boldsymbol{x}-\boldsymbol{y}\|$；函数 f 称为 μ-强凸（$\mu>0$），若对任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}\,f$ 和 $\xi(\boldsymbol{x})\in\partial f(\boldsymbol{x})$ 都有 $f(\boldsymbol{y})\geqslant f(\boldsymbol{x})+\langle\xi(\boldsymbol{x}),\boldsymbol{y}-\boldsymbol{x}\rangle+\frac{\mu}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2$。

定义 5[14]　（KL 性质）函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 为正常下半连续函数，令 $-\infty<\eta_1<\eta_2\leqslant+\infty$，记 $[\eta_1<f<\eta_2]:=\{\boldsymbol{x}\in\mathbf{R}^n:\eta_1<f(\boldsymbol{x})<\eta_2\}$，称函数 f 在 $\boldsymbol{x}^*\in\mathrm{dom}\,f$ 有 KL 性质，若存在 $\eta\in(0,+\infty]$，$\boldsymbol{x}^*$ 的某邻域 U 及连续的凹函数 $\varphi:[0,\eta)\to\mathbf{R}_+$，使得：

1）$\varphi(0)=0$；

2）φ 在 $(0,\eta)$ 上是连续可微的，并且 φ 在 0 处也是连续的；

3）对 $\forall\,\boldsymbol{s}\in(0,\eta)$，$\varphi'(\boldsymbol{s})>0$；

4）对 $\forall\,\boldsymbol{x}\in U\cap[f(\boldsymbol{x}^*)<f<f(\boldsymbol{x}^*)+\eta]$，KL 不等式都成立：$\varphi'(f(\boldsymbol{x})-f(\boldsymbol{x}^*))\mathrm{dist}(0,\partial f(\boldsymbol{x}))\geqslant 1$。

若 f 在定义域 dom f 的每一点处都满足 KL 性质，则称 f 为 KL 函数。

引理 1[14]　（一致 KL 性质）假设 Ω 为紧集，函数 $f:\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 为正常下半连续函数。若函数 f 在 Ω 上为一个常值并且在 Ω 的每一点都满足 KL 性质，则存在 $\varepsilon>0$，$\eta>0$，$\varphi\in\xi_\eta$，使得对任意 $\bar{\boldsymbol{x}}\in\Omega$ 及任意 $\boldsymbol{x}$ 属于如下交集 $\{\boldsymbol{x}\in\mathbf{R}^n:\mathrm{dist}(\boldsymbol{x},\Omega)<\varepsilon\}\cap\{f(\bar{\boldsymbol{x}})<f<f(\bar{\boldsymbol{x}})+\eta\}$，有 $\varphi'(f(\boldsymbol{x})-f(\bar{\boldsymbol{x}}))\mathrm{dist}(0,\partial f(\boldsymbol{x}))\geqslant 1$ 成立。

在一些实际应用中，许多函数都满足 KL 性质，如半代数函数、实解析函数、次解析函数、强凸函数等[10]。Bregman 距离首次在文献[15]中引入，在各种优化算法中起着重要的作用。作为平方欧氏距离的推广，Bregman 距离与欧氏距离有许多相似的优良性状。然而，Bregman 距离不是度量，因为它既不满足三角不等式，也不满足对称性。对一个可微凸函数 ξ，与它相关的 Bregman 距离的定义为：$\Delta_\xi(\boldsymbol{x},\boldsymbol{y})=\xi(\boldsymbol{x})-\xi(\boldsymbol{y})-\langle\nabla\xi(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle$。特别的，如果令上式中的 $\xi(\boldsymbol{x}):=\|\boldsymbol{x}\|^2$，则上式可简化为 $\|\boldsymbol{x}-\boldsymbol{y}\|^2$，即是经典的欧氏距离。下面将列举一些与本文相关的 Bregman 距离的性质。

命题 2　令 ξ 是一个可微凸函数，$\Delta_\xi(\boldsymbol{x},\boldsymbol{y})$ 是与它相关的 Bregman 距离。

1）非负性：对任意的 $\boldsymbol{x},\boldsymbol{y}$，有 $\Delta_\xi(\boldsymbol{x},\boldsymbol{y})\geqslant 0$，$\Delta_\xi(\boldsymbol{x},\boldsymbol{x})=0$；

2）凸性：$\Delta_\xi(\boldsymbol{x},\boldsymbol{y})$ 关于自变量 $\boldsymbol{x}$ 是凸的，关于自变量 $\boldsymbol{y}$ 不一定是凸的；

3) 强凸性：如果函数 ξ 是 δ-强凸的，那么对任意的 $\boldsymbol{x},\boldsymbol{y}$，都有 $\Delta_\xi(\boldsymbol{x},\boldsymbol{y})\geqslant\frac{\delta}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$。

2　收敛性分析

在建立算法 Majorized BADMM 的收敛性之前需做如下假设。

假设 1　1) 目标函数 $f:\mathbf{R}^n\to\mathbf{R}$ 是梯度 Lipschitz 连续的凸函数，其梯度的 Lipschitz 常数为 l_f，$g:\mathbf{R}^m\to\mathbf{R}\cup\{+\infty\}$ 是正常下半连续函数，$\lambda_{\max}$ 表示半正定矩阵 $\hat{\boldsymbol{\Sigma}}_f$ 的最大特征值；

2) 存在 $\mu_0>0$，使得对任意的 $\boldsymbol{x}\in\mathbf{R}^l$ 都有 $\mu_0\|\boldsymbol{x}\|^2\leqslant\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\|^2$，$\boldsymbol{B}$ 是单射；

3) ξ 是 μ_1 强凸的；

4) $\nabla\xi$ 和 $\nabla\psi$ 是 Lipschitz 连续的，它们的 Lipschitz 常数分别为 l_ξ,l_ψ；

5) 参数 β 满足 $\beta>\bar{\beta}=\frac{4[(l_f+\lambda_{\max})^2+(l_f+l_\xi+\lambda_{\max})^2]}{\mu_0\mu_1}$。

基于算法 Majorized BADMM 中每个子问题的最优性必要条件，有：

$$\begin{cases}0\in\partial g(\boldsymbol{y}^{k+1})-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\lambda}^k-\beta\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{B}\boldsymbol{y}^{k+1})+\nabla\psi(\boldsymbol{y}^{k+1})-\nabla\psi(\boldsymbol{y}^k),\\ 0=\nabla f(\boldsymbol{x}^k)+\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^k+\beta\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})+\nabla\xi(\boldsymbol{x}^{k+1})-\nabla\xi(\boldsymbol{x}^k),\\ \boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k+\beta(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})。\end{cases}\tag{5}$$

且始终假设迭代序列 $\{\boldsymbol{z}^k=(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k)\}$ 有界，令 $\{\hat{\boldsymbol{z}}^k=(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^{k-1},\boldsymbol{x}^{k-1})\}$，$\{\hat{\boldsymbol{z}}=(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{x}',\hat{\boldsymbol{x}})\}$，算法 Majorized BADMM 的收敛性分析需要基于如下效益函数：

$$\hat{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{x}',\hat{\boldsymbol{x}})=\hat{L}_\beta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{x}')+\frac{\sigma_0}{2}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2。\tag{6}$$

下面的引理 2 表明效益函数的单调递减性。

引理 2　若假设 1 成立，则存在 $\sigma_i>0,i=0,1$，使得 $\sigma_1\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2\leqslant\hat{L}(\hat{\boldsymbol{z}}^k)-\hat{L}(\hat{\boldsymbol{z}}^{k+1})$，其中 $\hat{L}$ 的定义为式(6)，且 $\sigma_0=\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}$，$\sigma_1=\frac{\mu_1}{2}-\frac{2(l_\xi+\lambda_{\max})^2}{\beta\mu_0}-\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}$。

证明　先证对任意的 $k\in\mathbf{N}$，有：

$$\|\lambda^{k+1}-\lambda^k\|^2\leqslant\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\mu_0}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2+\frac{2(l_\xi+\lambda_{\max})^2}{\mu_0}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2。\tag{7}$$

事实上，由式(5)中的第 2 个式子有：

$$\nabla f(\boldsymbol{x}^k)+\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^k+\beta\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})+\nabla\xi(\boldsymbol{x}^{k+1})-\nabla\xi(\boldsymbol{x}^k)=0,$$

结合算法 1 中步骤 4 可得：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^{k+1}=-\nabla f(\boldsymbol{x}^k)+\nabla\xi(\boldsymbol{x}^k)-\nabla\xi(\boldsymbol{x}^{k+1})-\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k),$$

令 $k=k-1$，有：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^k=-\nabla f(\boldsymbol{x}^{k-1})+\nabla\xi(\boldsymbol{x}^{k-1})-\nabla\xi(\boldsymbol{x}^k)-\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}),$$

$$\|\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k)\|^2\leqslant2(l_f+l_\xi+\lambda_{\max})^2\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2+2(l_\xi+\lambda_{\max})^2\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2,$$

由假设 1 的 2)可得 $\|\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k)\|^2\geqslant\mu_0\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2$，结合上式有：

$$\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2\leqslant\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\mu_0}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2+\frac{2(l_\xi+\lambda_{\max})^2}{\mu_0}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2。$$

即式(7)成立。接下来证明：

$$\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^{k+1},\boldsymbol{x}^k)-\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^{k-1})\leqslant-\frac{\mu_1}{2}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2+\frac{1}{\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2。\tag{8}$$

由算法 1 中关于 $\boldsymbol{y}$ 的子问题有：

$$\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)\leqslant\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^k)。\tag{9}$$

由假设 1 的 3)和命题 2 的 3)有：$\Delta_\xi(\boldsymbol{x}^{k+1},\boldsymbol{x}^k)\geqslant\frac{\mu_1}{2}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2$，结合算法 1 中关于 $\boldsymbol{x}$ 的子问题有：

$$\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)\leqslant\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)-\frac{\mu_1}{2}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2, \tag{10}$$

$$\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^{k+1},\boldsymbol{x}^k)-\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{x}^k)=\langle\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k,\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1}\rangle=\frac{1}{\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2, \tag{11}$$

由 $\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^k)\leqslant\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^{k-1})$,并结合式(9)、(10)、(11),可得式(8)成立。

最后,结合式(7)和式(8)可得:

$$\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^{k+1},\boldsymbol{x}^k)-\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^{k-1})\leqslant$$
$$\left(\frac{2(l_\xi+\lambda_{\max})^2}{\beta\mu_0}-\frac{\mu_1}{2}\right)\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2+\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2。$$

即

$$\hat{L}_\beta(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1},\boldsymbol{\lambda}^{k+1},\boldsymbol{x}^k)+\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2\leqslant$$
$$\hat{L}_\beta(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{\lambda}^k,\boldsymbol{x}^{k-1})+\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2-\left(\frac{\mu_1}{2}-\frac{2(l_\xi+\lambda_{\max})^2}{\beta\mu_0}-\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}\right)\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2。$$

令 $\sigma_0=\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}$,$\sigma_1=\frac{\mu_1}{2}-\frac{2(l_\xi+\lambda_{\max})^2}{\beta\mu_0}-\frac{2(l_f+l_\xi+\lambda_{\max})^2}{\beta\mu_0}$,由假设1的5)显然有 σ_i 都是正数,因此引理2成立。 证毕

下面的定理1将得到算法 Majorized BADMM 的全局收敛性和序列$\{\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|\}$是渐进正则的。

定理1 若假设成立且由算法 Majorized BADMM 产生的迭代序列$\{\boldsymbol{z}^k:=(\boldsymbol{x}^k,\boldsymbol{y}^k,\boldsymbol{z}^k)\}$有界,则:

1) $\sum_{k=0}^{\infty}\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|^2<+\infty$,且有序列$\{\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|\}$是渐进正则的,即当 $k\to\infty$时,$\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|\to0$;

2) 序列$\{\boldsymbol{z}^k\}$的任意一个聚点都是 L_β 的一个稳定点。

证明 由于$\{\boldsymbol{z}^k\}$是有界的,因此$\{\hat{\boldsymbol{z}}^k\}$也有界且至少存在一个聚点,则存在子序列$\{\hat{\boldsymbol{z}}^{k_j}\}$,s.t. $\lim_{j\to+\infty}\hat{\boldsymbol{z}}^{k_j}=\hat{\boldsymbol{z}}^*$。因为 f 是连续可微的,g 是下半连续函数,所以 $\hat{L}(\cdot)$为下半连续函数,因此,$\liminf_{j\to\infty}\hat{L}(\hat{\boldsymbol{z}}^{k_j})\geqslant\hat{L}(\hat{\boldsymbol{z}}^*)$,则$\{\hat{L}(\hat{\boldsymbol{z}}^{k_j})\}$有下界。由引理2知$\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$单调递减,进而$\{\hat{L}(\hat{\boldsymbol{z}}^{k_j})\}$收敛。因此,$\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$单调递减且有收敛子列。于是$\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$整列收敛且对任意的 $k\in\boldsymbol{Z}$,都有 $\hat{L}(\hat{\boldsymbol{z}}^k)\geqslant\hat{L}(\hat{\boldsymbol{z}}^*)$。固定 $k\in\mathbf{N}$,由引理2可知:

$$\sigma_1\sum_{i=0}^{k}\|\boldsymbol{x}^i-\boldsymbol{x}^{i+1}\|^2\leqslant\sum_{i=0}^{k}(\hat{L}(\hat{\boldsymbol{z}}^i)-\hat{L}(\hat{\boldsymbol{z}}^{i+1}))=\hat{L}(\hat{\boldsymbol{z}}^0)-\hat{L}(\hat{\boldsymbol{z}}^{k+1})\leqslant\hat{L}(\hat{\boldsymbol{z}}^0)-\hat{L}(\hat{\boldsymbol{z}}^*)<+\infty,$$

因为 k 的任意性,所以 $\sum_{k=0}^{\infty}\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|^2<+\infty$。结合式(7)可知 $\sum_{k=0}^{\infty}\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\|^2<+\infty$。因为 $\boldsymbol{B}$ 是单射,则显然会存在 $\mu_{\boldsymbol{B}}>0$,使得:

$$\beta^2\mu_{\boldsymbol{B}}\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|^2\leqslant\|\beta\boldsymbol{B}(\boldsymbol{y}^k-\boldsymbol{y}^{k+1})\|^2=\|\boldsymbol{\lambda}^{k-1}-\boldsymbol{\lambda}^k+\beta\boldsymbol{A}\boldsymbol{x}^k+\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k-\beta\boldsymbol{A}\boldsymbol{x}^{k+1}\|^2\leqslant$$
$$3(\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\|^2+\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\|^2+\beta^2\|\boldsymbol{A}\|^2\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2), \tag{12}$$

则 $\sum_{k=0}^{\infty}\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|^2<+\infty$。因此 $\sum_{k=0}^{\infty}\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|^2<+\infty$,且当 $k\to\infty$时有$\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|\to0$。设 $\boldsymbol{z}^*=(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*)$是序列$\{\boldsymbol{z}^k\}$的任意一个聚点,设序列$\{\boldsymbol{z}^{k_j}\}$是序列$\{\boldsymbol{z}^k\}$的一个子列,且收敛到 $\boldsymbol{z}^*$。因为$\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|\to0(k\to\infty)$,则序列$\{\boldsymbol{z}^{k_j}\}$和序列$\{\boldsymbol{z}^{k_j+1}\}$有相同的极限点 $\boldsymbol{z}^*$。因为序列$\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$收敛,所以有序列$\{g(\boldsymbol{y}^k)\}$收敛。结合式(5)和乘子的迭代式可得:

$$\lambda^{k+1}=\lambda^k+\beta(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1}),$$
$$-\nabla f(\boldsymbol{x}^k)=\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^{k+1}+\nabla\xi(\boldsymbol{x}^{k+1})-\nabla\xi(\boldsymbol{x}^k),$$
$$\partial g(\boldsymbol{y}^{k+1})\ni\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\lambda}^k+\beta\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{B}\boldsymbol{y}^{k+1})+\nabla\psi(\boldsymbol{y}^k)-\nabla\psi(\boldsymbol{y}^{k+1})=$$
$$\boldsymbol{B}^{\mathrm{T}}\lambda^{k+1}+\beta\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{A}\boldsymbol{x}^{k+1})+\nabla\psi(\boldsymbol{y}^k)-\nabla\psi(\boldsymbol{y}^{k+1})。$$

将上式中的所有 k 换成 k_j,并令 $j\to+\infty$,则有:$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^*=-\nabla f(\boldsymbol{x}^*)$,$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\lambda}^*\in\partial g(\boldsymbol{y}^*)$,$\boldsymbol{A}\boldsymbol{x}^*=\boldsymbol{B}\boldsymbol{y}^*$。

因此,由命题1的3)和定义3可知,$\boldsymbol{z}^*$ 是 L_β 的一个稳定点。 证毕

接下来将提出对于证明强收敛性非常重要的一个刻画相对误差性的引理。

引理3 若假设1成立,则存在 $\kappa>0$ 使得对任意的 k,有:

$$\mathrm{dist}(0,\partial\hat{L}(\hat{\boldsymbol{z}}^{k+1}))\leqslant\kappa(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|+\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|)。$$

证明 由 $\hat{L}$ 的定义和式子:$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^{k+1}=-\nabla f(\boldsymbol{x}^k)+\nabla\xi(\boldsymbol{x}^k)-\nabla\xi(\boldsymbol{x}^{k+1})-\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)$ 可得:

$$\begin{aligned}\partial\hat{L}_x(\hat{\boldsymbol{z}}^{k+1})&=\nabla f(\boldsymbol{x}^k)+\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^{k+1}+\beta\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1})+\sigma_0(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)=\\&\nabla\xi(\boldsymbol{x}^k)-\nabla\xi(\boldsymbol{x}^{k+1})+\sigma_0(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)+\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k),\end{aligned}$$

由式(5)的第1个式子有:$0\in\partial g(\boldsymbol{y}^{k+1})-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\lambda}^{k+1}-\beta\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{A}\boldsymbol{x}^{k+1})+\nabla\psi(\boldsymbol{y}^{k+1})-\nabla\psi(\boldsymbol{y}^k)$,则有:

$$\partial\hat{L}_y(\hat{\boldsymbol{z}}^{k+1})=\nabla\psi(\boldsymbol{y}^k)-\nabla\psi(\boldsymbol{y}^{k+1})+\beta\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{A}\boldsymbol{x}^{k+1})+\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}),$$

显然有 $\partial\hat{L}_{\hat{x}}(\hat{\boldsymbol{z}}^{k+1})=-\sigma_0(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)$,$\partial\hat{L}_\lambda(\hat{\boldsymbol{z}}^{k+1})=\boldsymbol{A}\boldsymbol{x}^{k+1}-\boldsymbol{B}\boldsymbol{y}^{k+1}=\dfrac{1}{\beta}(\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k)$ 成立,并且有 $\partial\hat{L}_{x'}(\hat{\boldsymbol{z}}^{k+1})=\hat{\boldsymbol{\Sigma}}_f(\boldsymbol{x}^k-\boldsymbol{x}^{k+1})$,使得

$$\mathrm{dist}(0,\partial\hat{L}(\hat{\boldsymbol{z}}^{k+1}))\leqslant\kappa_0(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|+\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|+\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|),\kappa_0>0。$$

由式(7)可得:

$$\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|\leqslant\frac{\sqrt{2}(l_f+l_\xi+\lambda_{\max})}{\sqrt{\mu_0}}(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|)=\kappa_1(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|),\tag{13}$$

其中:$\kappa_1=\dfrac{\sqrt{2}(l_f+l_\xi+\lambda_{\max})}{\sqrt{\mu_0}}$,又由式(12)可得:

$$\begin{aligned}\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|&\leqslant\frac{\sqrt{3}}{\beta\sqrt{\mu_{\boldsymbol{B}}}}((\beta\|\boldsymbol{A}\|+\kappa_1)\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|+2\kappa_1\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\kappa_1\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|)=\\&\kappa_2(\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|+\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|),\end{aligned}\tag{14}$$

其中:$\kappa_2=\dfrac{\sqrt{3}(2\kappa_1+\beta\|\boldsymbol{A}\|)}{\beta\sqrt{\mu_{\boldsymbol{B}}}}$。令 $\kappa=\kappa_0(\kappa_1+\kappa_2+1)$,则有:

$$\mathrm{dist}(0,\partial\hat{L}(\hat{\boldsymbol{z}}^{k+1}))\leqslant\kappa(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|+\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|)。$$

因此引理得证。 证毕

下面利用引理2、定理1以及引理3中的相关结论建立算法 Majorized BADMM 的强收敛性。

定理2 令 Ω 为序列 $\{\hat{\boldsymbol{z}}^k\}$ 的聚点集,若假设成立,序列 $\{\boldsymbol{z}^k\}$ 有界且 $\hat{L}(\cdot)$ 在 Ω 上满足 KL 性质,则有:1) $\sum\limits_{k=0}^{\infty}\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|<+\infty$;2) 序列 $\{\boldsymbol{z}^k\}$ 收敛到问题(1)的一个稳定点。

证明 由定理1可知序列 $\{\boldsymbol{x}^k\}$ 是渐近正则的,则序列 $\{\boldsymbol{x}^k\}$ 和 $\{\boldsymbol{x}^{k+1}\}$ 有相同的聚点。令 $\hat{\boldsymbol{z}}^*=(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{\lambda}^*,\boldsymbol{x}^*,\boldsymbol{x}^*)\in\Omega$,再令序列 $\{\hat{\boldsymbol{z}}^{k_j}\}$ 为序列 $\{\hat{\boldsymbol{z}}^k\}$ 的一个收敛子列,且收敛到 $\hat{\boldsymbol{z}}^*$。因为 f 是连续可微的,g 是下半连续函数,所以 $\hat{L}(\cdot)$ 为下半连续,则 $\hat{L}(\hat{\boldsymbol{z}}^{k_j})\to\hat{L}(\hat{\boldsymbol{z}}^*)$。因为由定理1可知序列 $\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$ 是收敛的,所以 $\hat{L}(\hat{\boldsymbol{z}}^k)\to\hat{L}(\hat{\boldsymbol{z}}^*)$。因此函数 $\hat{L}(\cdot)$ 在集合 Ω 上是一个常值。分2种情况讨论 $\hat{L}(\hat{\boldsymbol{z}}^k)$。

情形1,存在 $k_0\in\mathbf{N}$ 使得 $\hat{L}_{k_0}=\hat{L}(\hat{\boldsymbol{z}}^*)$,则由引理2和序列 $\{\hat{L}(\hat{\boldsymbol{z}}^k)\}$ 是单调非增的可知对 $\forall k>k_0$,有:

$$\sigma_1\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2\leqslant\hat{L}(\hat{\boldsymbol{z}}^k)-\hat{L}(\hat{\boldsymbol{z}}^{k+1})\leqslant\hat{L}(\hat{\boldsymbol{z}}^{k_0})-\hat{L}(\hat{\boldsymbol{z}}^*)=0,$$

结合式(13)和(14)可知序列 $\{\boldsymbol{z}^k\}$ 除去有限项后是一个常数列,因此序列 $\{\boldsymbol{z}^k\}$ 是收敛的。

情形2,对 $\forall k\in\mathbf{N}$,有 $\hat{L}(\hat{\boldsymbol{z}}^k)\geqslant\hat{L}(\hat{\boldsymbol{z}}^*)$。由假设 $\hat{L}(\cdot)$ 满足 KL 不等式,结合引理1可知存在 $\eta>0,\delta>0$,$\varphi\in\varphi_\eta$,使得对所有满足 $\mathrm{dist}(\hat{\boldsymbol{z}},\Omega)<\delta$ 和 $\hat{L}(\hat{\boldsymbol{z}}^*)<\hat{L}(\hat{\boldsymbol{z}})<\hat{L}(\hat{\boldsymbol{z}}^*)+\eta$ 的 $\hat{\boldsymbol{z}}$ 有下式成立:

$$\varphi'(\hat{L}(\hat{\boldsymbol{z}})-\hat{L}(\hat{\boldsymbol{z}}^*))\mathrm{dist}(0,\partial\hat{L}(\hat{\boldsymbol{z}}))\geqslant 1。$$

由 Ω 的定义知 $\lim\limits_{k\to\infty}\mathrm{dist}(\hat{\boldsymbol{z}}^k,\Omega)=0$。结合 $\hat{L}(\hat{\boldsymbol{z}}^k)\to\hat{L}(\hat{\boldsymbol{z}}^*)$ 可得:存在 $k_1\in\mathbf{N}$,使得对任意的 $k\geqslant k_1$,都有

$\mathrm{dist}(\hat{z}^k,\Omega)<\delta$ 和 $\hat{L}(\hat{z}^k)<\hat{L}(\hat{z}^*)+\eta$。令 $k>k_1$,则:

$$\hat{z}^k\in\{\hat{z}:\mathrm{dist}(\hat{z},\Omega)<\delta\}\cap\{\hat{z}:\hat{L}(\hat{z}^*)<\hat{L}(\hat{z})<\hat{L}(\hat{z}^*)+\eta\},$$

因此有 $\mathrm{dist}(0,\partial\hat{L}(\hat{z}^k))\varphi'(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))\geqslant 1$。由引理 3 可得:

$$\frac{1}{\varphi'(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))}\leqslant\kappa(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|),$$

再由 φ 的凹性可得:

$$\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^{k+1})\leqslant\kappa(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|)\times$$
$$[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))],$$

结合引理 2 有:

$$\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|^2\leqslant\frac{\kappa}{\sigma_1}(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|)\times$$
$$[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))],$$
$$4\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|\leqslant 2(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|)^{1/2}\times$$
$$2\sqrt{\frac{\kappa}{\sigma_1}}[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))]^{1/2},$$

由不等式 $2ab\leqslant a^2+b^2$,有:

$$2(\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|)^{1/2}\times 2\sqrt{\frac{\kappa}{\sigma_1}}[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))]^{1/2}\leqslant$$
$$\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|+4\frac{\kappa}{\sigma_1}[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))],$$

所以:

$$4\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^k\|\leqslant\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|+\|\boldsymbol{x}^{k-1}-\boldsymbol{x}^{k-2}\|+\|\boldsymbol{x}^{k-2}-\boldsymbol{x}^{k-3}\|+$$
$$4\frac{\kappa}{\sigma_1}[\varphi(\hat{L}(\hat{z}^k)-\hat{L}(\hat{z}^*))-\varphi(\hat{L}(\hat{z}^{k+1})-\hat{L}(\hat{z}^*))],$$

等价于:

$$\sum_{i=k_1}^{k}\|\boldsymbol{x}^{i+1}-\boldsymbol{x}^i\|\leqslant 3\|\boldsymbol{x}^{k_1}-\boldsymbol{x}^{k_1-1}\|+2\|\boldsymbol{x}^{k_1-1}-\boldsymbol{x}^{k_1-2}\|+\|\boldsymbol{x}^{k_1-2}-\boldsymbol{x}^{k_1-3}\|+4\frac{\kappa}{\sigma_1}\varphi(\hat{L}(\hat{z}^{k_1})-\hat{L}(\hat{z}^*))。$$

因为 k 的任意性,所以有 $\sum_{k=0}^{\infty}\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|<+\infty$。又由引理 3 中的式(13)、(14)可知:

$$\sum_{k=0}^{\infty}(\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|+\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|)<+\infty,$$

因为 $\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|=(\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|^2+\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|^2+\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\|^2)^{1/2}\leqslant\|\boldsymbol{x}^k-\boldsymbol{x}^{k+1}\|+\|\boldsymbol{y}^k-\boldsymbol{y}^{k+1}\|+\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k+1}\|$,所以 $\sum_{k=0}^{\infty}\|\boldsymbol{z}^k-\boldsymbol{z}^{k+1}\|<+\infty$,则序列$\{\boldsymbol{z}^k\}$是一个柯西序列,因此序列$\{\boldsymbol{z}^k\}$收敛,结合定理 1 证明完成。　　证毕

3　数值实验

本节将提出的算法 Majorized BADMM 应用于解决稀疏逻辑回归问题[17],并将相应的数值表现与经典式(2)进行比较。

一般的稀疏逻辑回归问题的形式如下:

$$\min_x\sum_{i=1}^{n}\log(1+\exp(-b_i\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}))+n\rho\|\boldsymbol{x}\|_q。\tag{15}$$

其中:$\boldsymbol{a}_i\in\mathbf{R}^d$ 是样本 $i(i=1,2,\cdots,n)$的特征向量,$b_i\in\{1,-1\}$是相应的二分类,ρ 是一个正则化参数,$\|\cdot\|_q$ 是 l_q 范数,对任意的 $\boldsymbol{x}\in\mathbf{R}^d$ 有 $\|\boldsymbol{x}\|_q=\left(\sum_{i=1}^{n}|x_i|^q\right)^{1/q}$, $q\in(0,1]$。令 $\mathbf{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)^{\mathrm{T}}\in\mathbf{R}^{n\times d}$,$\boldsymbol{b}=$

$(b_1, b_2, \cdots, b_n)^{\mathrm{T}} \in \mathbf{R}^n$,则$(\boldsymbol{a}_i, b_i)(i \in \{1,2,\cdots,n\})$是训练集$(\boldsymbol{a}, \boldsymbol{b})$的一个样本。该问题有 n 个样本,维数为 d。

对于 ADMM 和 Majorized BADMM,需要将式(15)重新表述为以下等价形式:

$$\min \sum_{i=1}^{n} \log(1+\exp(-b_i \boldsymbol{a}_i^{\mathrm{T}} \boldsymbol{x})) + n\rho \|\boldsymbol{y}\|_q,$$
$$\text{s.t. } \boldsymbol{x} - \boldsymbol{y} = 0。 \tag{16}$$

显然式(16)属于式(1)的形式,其中 $f(x)=\sum_{i=1}^{n} \log(1+\exp(-b_i \boldsymbol{a}_i^{\mathrm{T}} \boldsymbol{x}))$,$g(\boldsymbol{y})=n\rho\|\boldsymbol{y}\|_q$。易知$\nabla f$ 是 Lipschitz 连续的,Lipschitz 常数是 $L=\frac{1}{4}\|\boldsymbol{A}\|_2^2$。因为 f 是梯度 Lipschitz 连续的,所以存在一个半正定矩阵$\hat{\boldsymbol{\Sigma}}_f$ 使得对任意给定的 $\boldsymbol{x}' \in \mathbf{R}^d$,有:

$$f(\boldsymbol{x}) \leqslant \hat{f}(\boldsymbol{x}, \boldsymbol{x}') := f(\boldsymbol{x}') + \langle \nabla f(\boldsymbol{x}'), \boldsymbol{x}-\boldsymbol{x}' \rangle + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}'\|_{\hat{\boldsymbol{\Sigma}}_f}^2,$$

令 $C_i=-b_i \boldsymbol{a}_i^{\mathrm{T}}$,经简单的推算有:$\nabla f(\boldsymbol{x})=\sum_{i=1}^{n} \frac{\boldsymbol{C}_i e^{C_i x}}{1+e^{C_i x}}$,$\nabla^2 f(\boldsymbol{x})=\sum_{i=1}^{n} \boldsymbol{C}_i \boldsymbol{C}_i^{\mathrm{T}} \frac{e^{C_i x}}{(1+e^{C_i x})^2} \leq \frac{1}{4}\sum_{i=1}^{n} \boldsymbol{C}_i \boldsymbol{C}_i^{\mathrm{T}}$,因此,$\hat{\boldsymbol{\Sigma}}_f$ 近似项可以选择为$\hat{\boldsymbol{\Sigma}}_f=\frac{1}{4}\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}$,其中 $\boldsymbol{C}=[\boldsymbol{C}_1, \cdots, \boldsymbol{C}_n] \in \mathbf{R}^{d \times n}$。

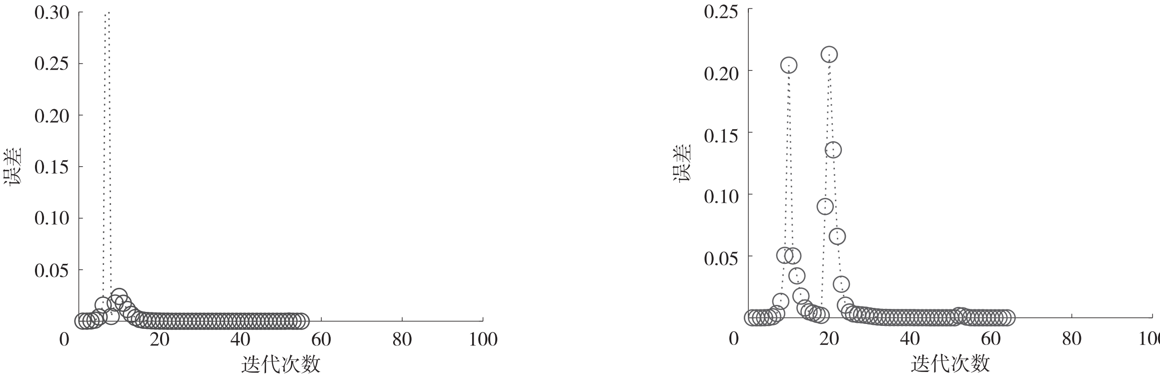
本文实验测试了 Majorized BADMM 和式(2)的数值表现,以解决问题(16)在 $q=1$ 和 $q=1/2$ 的2种不同情况。其中 $q=1/2$ 时,问题(16)是一个非凸优化问题。在实验中,选择来自 LIBSVM 数据库(https://www.csie.ntu.edu.tw/cjlin/libsvmtools/)数据 a2000a、a1a 来进行计算,。正则化参数 ρ 的选取标准为 $\rho=\frac{\gamma}{N}\|\boldsymbol{A}^{\mathrm{T}} b\|_\infty$,其中 N 是样本的维度大小,$0<\gamma<1$。当 $q=1$ 时,问题(16)是一个凸优化问题,选取式(2)的增广拉格朗日函数的惩罚参数 $\beta=1$。当 $q=1/2$ 时问题(16)是一个非凸优化问题,选取式(2)的增广拉格朗日函数的惩罚参数为 $\beta=3L$,满足文献[20]中的假设 $\beta>2L$。Majorized BADMM 的增广拉格朗日函数的惩罚参数 β 按照文献[16]假设中同样的方式来选取。最初令 $\beta=0.1\bar{\beta}$,其中 $\bar{\beta}$ 由假设1的5)给出。在第 k 次迭代,计算 $n_k=\|\boldsymbol{x}^k\|^2+\|\boldsymbol{y}^k\|^2$,$s_k=\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|^2+\|\boldsymbol{y}^k-\boldsymbol{y}^{k-1}\|^2$。从第2次迭代开始,更新 $\beta=1.05\beta$,当 $\beta \geqslant 1.01\bar{\beta}$ 或序列满足 $n_k \geqslant 10^{10}$ 或 $s_k \geqslant 0.99 \cdot s_{k-1}$ 时,停止更新 β。运用 Majorized BADMM 和式(2)来解决问题(16)时都使用了与文献[1]相同的牛顿法来近似估算关于 $\boldsymbol{x}$ 的子问题。所有的方法都选择 $x_0=0$ 为初始点,终止条件都设置为 $e_k := \frac{\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|}{\max\{\|x^k\|, 1\}} \leqslant 1 \times 10^{-6}$。最大迭代次数设为5 000,所有的测试均在 MATLAB 2019b 上执行,运行的 PC 配置为 Windows 10 系统,Intel(R)Core(TM)i7-6500UCPU@2.50 GHz,8.00 GB 内存。

分别利用 Majorized BADMM 和 ADMM 求解问题(16),并选取2个不同维度的训练集进行测试。表1是 $q=1$ 和 $q=1/2$ 时2个算法的主要数值结果,可以看出2种算法都可以在几乎相同的目标函数值下实现最优解,并且 Majorized BADMM 比 ADMM 花费更少的计算时间。此外,为进一步观察测试算法的收敛性,绘制了 $q=1$ 和 $q=1/2$ 时误差 e_k 相对于迭代次数的演化曲线(图1和图2);再结合表1可以发现 Majorized BADMM 是收敛的且比 ADMM 更快地获得最优解,特别是随着样本维度的增大,Majorized BADMM 的优势更加突出。

表1 Majorized BADMM 和 ADMM 的数值比较结果

Tab.1 Numerical comparisons between Majorized BADMM and ADMM

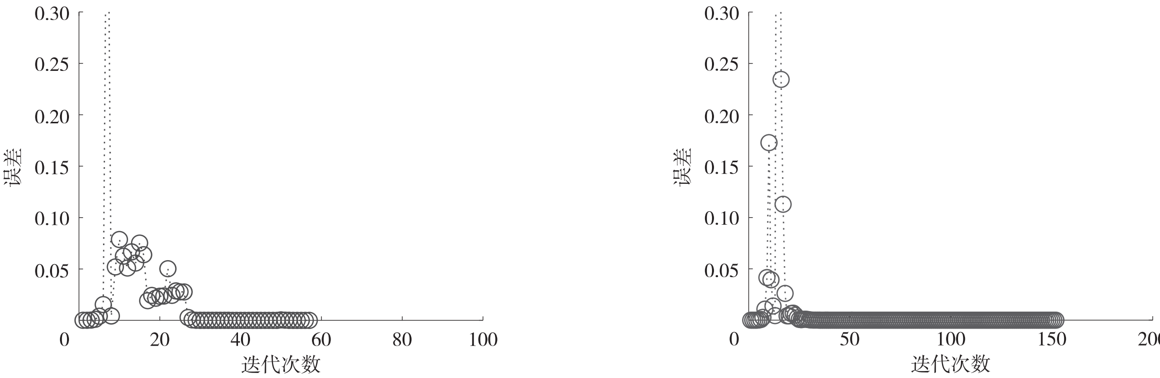
数据	Q	维度大小	Majorized BADMM			ADMM		
			迭代次数	CPY 计算时间/s	目标函数值	迭代次数	CPY 计算时间/s	目标函数值
a2000a	1	999×60	55	3.14	616.02	119	3.60	613.91
a2000a	1/2	999×60	57	3.03	672.07	228	7.43	670.5
a1a	1	1 604×118	64	15.46	357.55	147	21.18	355.08
a1a	1/2	1 604×118	152	10.06	742.52	395	36.76	740.5



a　a2000a　　　　b　a1a

图 1　当 $q=1$ 时,误差关于迭代次数的演化曲线

Fig. 1　Evolution curve of error with respect to the number of iterations when $q=1$



a　a2000a　　　　b　a1a

图 2　当 $q=1/2$ 时,误差关于迭代次数的演化曲线

Fig. 2　Evolution curve of error with respect to the number of iterations when $q=1/2$

4　结论和展望

本文基于一类两块可分非凸优化问题,提出了一类 Majorized Bregman 交替方向乘子法,新算法结合了 Majorized ADMM 和 Bregman ADMM 的优点。首先,在适当的假设条件下,建立了 Majorized BADMM 的全局收敛性,即迭代序列$\{z^k\}$的任意聚点都是 $L_\beta(\cdot)$的稳定点;其次,在效益函数满足 KL 性质的条件下,建立了 Majorized BADMM 的强收敛性,即迭代序列$\{z^k\}$整列收敛到 $L_\beta(\cdot)$的稳定点;最后,通过数值实验说明 Majorized BADMM 的有效性和优越性。

实际上,在本文基础上还有许多问题值得进一步讨论,比如:在较弱的条件下给出 Majorized BADMM 的线性收敛率;探究多分块的 Majorized BADMM;寻求更多关于非凸优化问题的实际算例,使 Majorized BADMM 的优势体现的更为充分;尝试将该方法推广到随机环境中,并应用于解决实际应用中出现的一些问题。

参考文献:

[1] BOYD S,PARIKH N,CHU E,PELEATO B,et al. Distributed optimization and statisticallearning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning,2010,3(1):1-112.

[2] GABAY D,MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation [J]. Computers & Mathematics with Applications,1976,2(1):17-40.

[3] LIONS P L, MERCIER B. Splitting algorithms for the sum of two nonlinear operators[J]. SIAM Journal on Numerical Analysis,1979,16(6):964-979.

[4] LI G Y,PONG T K. Global convergence of splitting methods for nonconvex composite optimization[J]. SIAM Journal on Optimization,2015,25(4):2434-2460.

[5] HONG M Y,LUO Z Q,RAZAVIYAYN M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems[J]. SIAM Journal on Optimization,2016,26(1):337-364.

[6] GUO K,HAN D R,WU T. Convergence of alternating direction method for minimizing sum of two nonconvex functions with Linear constraints[J]. International Journal of Computer Mathematics,2017,12(5):1139-1162.

[7] WANG Y,YIN W T,ZENG J S. Global convergence of admm in nonconvex nonsmooth optimization[J]. Journal of Scientific Computing,2019,78(1):29-63.

[8] BOT R I, NGUYEN D K. The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates[J]. Mathematics of Operations Research,2020,45(2):682-712.

[9] HAN D R, SUN D F, ZHANG L W. Linear rate convergence of the alternating direction method of multipliers for convex composite programming[J]. Mathematics of Operations Research,2017,43(2):622-637.

[10] WANG F H,CAO W F,XU Z B. Convergence of multi-block bregmanadmm for nonconvexcomposite problems[J]. Science China Information Sciences,2018,61(12):1-12.

[11] LI M,SUN D F,TOH K C. A majorized admm with indefinite proximal terms for linearly constrained convex composite optimization[J]. SIAM Journal on Optimization,2016,26(2):922-950.

[12] ZHANG N,WU J,ZHANG L W. A linearly convergent majorized admm with indefinite proximalterms for convex composite programming and its applications[J]. Mathematics of Computation,2020,89(324):1867-1894.

[13] ROCKEFELLERSR T,WETS R J B. Variational analysis[M]. Berlin:Springer Science and Business Media,2009.

[14] BOLTE J, DANIILIDIS A, LEWISA. The Lojasiewicz inequality for nonsmoothsubanalytic functions with applications to subgradient dynamical systems[J]. SIAM Journal on Optimization,2006,17(4):1205-1223.

[15] BREGMAN L M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming[J]. USSR Computational Mathematics and Mathematical Physics,1967,7(3):200-217.

[16] YANG L, PONG T K, CHEN X J. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction[J]. SIAM Journal on Imaging Sciences,2017,10(1):74-110.

[17] YUAN G X, HO C H, LIN C J. An improved glmnet forL1-regularized logistic regression[J]. Journal of Machine Learning Research,2012,13(64):1999-2030.

Operations Research and Cybernetics

Majorized Bregman Alternating Direction Method of Multipliers for Solving Nonconvex Two Block Optimization Problems

CHEN Jianhua, PENG Jianwen, LUO Honglin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: For a class of two block nonconvex optimization problems, a majorized alternating direction method of multipliers with Bregman distance is proposed. In order to make the subproblem of the problem easier to solve, maximizing the smooth term in the objective function with linear processing and a Bergman distance is added to the x-subproblem and the y-subproblem at the same time. Under appropriate assumptions, the global convergence of the algorithm is established. Secondly, when the benefit function satisfies the KL property, the strong convergence of the algorithm is established. Numerical experiments are carried out on the algorithm, and the results show that the algorithm is an effective method.

Keywords: alternating direction method of multipliers; Bregman distance; nonconvex optimization problem; KL property; convergence

(责任编辑 陈 乔)